I. Troch, F. Breitenecker, eds. ISBN 978-3-901608-35-3

CPL CLUSTERING BASED ON LINEAR DEPENDENCIES

Leon Bobrowski
Białystok Technical University, Poland
Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

Corresponding author: L. Bobrowski
Faculty of Computer Science, Białystok Technical University
ul. Wiejska 45A, 15-351 Białystok, Poland
E-mail: leon@ibib.waw.pl

Abstract: Discovering linear dependencies in data sets is discussed in the paper as a part of data mining approach [1]. The proposed method is based on the minimization of a special type of convex and piecewise linear (*CPL*) criterion functions defined on a given data set C [2]. The division of the set C into a family of linearly dependent clusters C_k allows to form a family of local regression type models. As a result, each subset C_k can be characterized C_k by its own linear model. The *K-plans* algorithm which is similar to the *K-means* algorithm can be used for dividing the set C into a family of linearly dependent clusters C_k. Also a different approach to this problem, based on the CPL criterion functions is discussed here.

1. Introduction

Data mining is a process of extracting hidden patterns from data [1]. Generally, data mining techniques can be useful in transformation of data sets into needed information. Such techniques are commonly used in a wide range of applications, such as marketing, fraud detection and scientific discovery. The term *patterns* could stand for regularities, trends, association rules or clusters in the explored data set.

A fundamental role in the cluster analysis is played by the *K-means* algorithm [2]. The *K-means* algorithm can be used for the purpose of dividing set C into a family of a priori given number K of clusters C_k ($k = 1,\ldots, K$). The central points $\mathbf{w}_k$ can be identified for each subset C_k during the *K-means* procedure through thee minimization of convex and piecewise linear (*CPL*) criterion functions [3]. Modification of the *K-means* algorithm into the *K-plans* algorithm has been proposed recently [4]. The proposed method is based on the minimization of a special type of the *CPL* criterion functions defined on a given data set C [2]. The basis exchange algorithms which are similar to linear programming allow one to find the minimum of these *CPL* function efficiently, even in the case of large multidimensional data sets [3]. The minimization of the *CPL* criterion function during the *K-plans* procedure allows to identify the actual values of the parameters w_k and θ_k of the central hyperplane $H(\mathbf{w}_k,\theta_k) = \{\mathbf{x}[n]\colon \ \mathbf{w}_k[n]^T\mathbf{x}[n] = \theta_k\}$ for each subset C_k. In the next step of the *K-plans* algorithm the division of the set C into the subsets C_k is modified and adopted to the actual central hyperplane $H(w_k,\theta_k)$. The central hyperplane $H(w_k,\theta_k)$ defines the local, linear dependency characteristic for a given subset C_k. As a result, each subset C_k can be characterized by its own linear model of dependencies.

The procedure of hidden linear dependencies extraction from data set different from the *K-plans* is also described and analyzed in the presented paper. The presented procedure is based on monotonicity properties of the the *CPL* criterion function. This procedure allows to identify a family of K subsets C_k and local, linear dependencies without assuming a priori the value of the number K. The number K of linear models results from a structure of the explored data set C.

The proposed approach can be used for solving a variety of data mining problems. One of them is discovering and analysing linearly dependent patterns (*models*) in data sets Data aggregation into linearly dependent subsets C_k can be combined in this approach with feature selection.

2. Feature vectors and central points

Let us take into considerations the set C of m feature vectors $\mathbf{x}_j[n] = [x_{j1},\ldots,x_{jn}]^T$ belonging to a given n-dimensional feature space $F[n]$ ($\mathbf{x}_j[n] \in F[n]$):

$$C = \{\mathbf{x}_j[n]\}, \ \ where \ \ j = 1,..., m \tag{1}$$

Components x_{ji} of the vector $\mathbf{x}_j[n]$ could be the numerical results of n standardized examinations of given objects O_j ($x_{ji} \in \{0,1\}$ or $x_{ji} \in R$). Each vector $\mathbf{x}_j[n]$ can be treated as a point of the n-dimensional feature space $F[n]$.

In accordance with the *K-means* algorithm, feature vectors $\mathbf{x}_j[n]$ are divided into subsets C_k on the basis of actual central points (*means*) $\mathbf{w}_k[n] = [w_1,\ldots,w_n]^T$ ($\mathbf{w}_k[n] \in R^n$):

$$(\forall j \in \{1,\ldots.., m\})\ (\forall k' \in \{1,\ldots.., K\})$$
$$\textit{if}\ (\rho(\mathbf{x}_j[n], \mathbf{w}_k[n]) < \rho(\mathbf{x}_j[n], \mathbf{w}_{k'}[n]),\ \textit{then}\ \ \mathbf{x}_j[n] \in C_k,\ \textit{and}\ j \in J_k \tag{2}$$

where $\rho(\mathbf{x}_j[n],\mathbf{w}_k[n])$ is the distance between the feature vector $\mathbf{x}_j[n]$ and the central point $\mathbf{w}_k[n]$, and J_k is the set of indices j of those vectors $\mathbf{x}_j[n]$ which have been allocated into the subset C_k.

The subsets C_k generated in accordance with the rule (2) allow to redefine new central points $\mathbf{w}_k'[n]$. The central points $\mathbf{w}_k'[n]$ of the subset C_k are computed through the minimization of the criterion function $Q_k(\mathbf{w}[n])$ defined on the elements $\mathbf{x}_j[n]$ of the subsets C_k:

$$Q_k(\mathbf{w}[n]) = \sum_{j \in J_k} \alpha_j \,\|\, \mathbf{x}_j[n] - \mathbf{w}[n] \| \tag{3}$$

where $\|\mathbf{x}_j[n] - \mathbf{w}[n]\|$ is the norm of the vector $\mathbf{x}_j[n] - \mathbf{w}[n]$ with the price α_j ($\alpha_j > 0$), and $\mathbf{w}_k'[n]$ is the minimum point of the function $Q_k(\mathbf{w}[n])$:

$$(\forall \mathbf{w}[n])\ \ Q_k(\mathbf{w}[n]) \geq Q_k(\mathbf{w}_k'[n]) \tag{4}$$

The new central points $\mathbf{w}_k'[n]$ allow to define new subsets C_k' in accordance with the rule (2). The *K-means* procedure stops when the difference between two successive central points $\mathbf{w}_k[n]$ and $\mathbf{w}_k'[n]$ is sufficiently small (in accordance with a given parameter $\varepsilon > 0$):

$$(\forall k \in \{1,\ldots.., K\})\quad \|\mathbf{w}_k'[n] - \mathbf{w}_k[n]\| \leq \varepsilon \tag{5}$$

The minimization procedure of the function $Q_k(\mathbf{w}[n])$ (3) depends on the choice of the norm $\|\, \mathbf{x}_j[n] - \mathbf{w}[n]\|$. Commonly used is the Euclidean norm L_2 [3]:

$$\|\mathbf{x}_j[n] - \mathbf{w}[n]\|_{L2} = (\mathbf{x}_j[n] - \mathbf{w}[n])^{\mathrm{T}}(\mathbf{x}_j[n] - \mathbf{w}[n])^{1/2} \tag{6}$$

The L_1 and L_∞ norms are also used in the *K-means* algorithm:

$$\|\, \mathbf{x}_j[n]\ - \mathbf{w}|[n]\,|_{L1} = \sum_{i = 1,\ldots,n} |\, \mathrm{x}_{ji} - \mathrm{w}_i \,| \tag{7}$$

and

$$\|\, \mathbf{x}_j[n]\ - \mathbf{w}[n]\|_{L\infty} = \max_i |\, \mathrm{x}_{ji} - \mathrm{w}_i \,| \tag{8}$$

where $\mathbf{w}[n] = [\mathrm{w}_1,\ldots\ldots.,\mathrm{w}_n]^{\mathrm{T}}$.

The minimum point $\mathbf{w}_k'[n]$ (4) of the function $Q_k(\mathbf{w}[n])$ (3) can be found analytically in the case the Euclidean norm L_2. The criterion function $Q_k(\mathbf{w}[n])$ (3) is convex and piecewise linear (*CPL*) in the case the norms L_1 (7) and L_∞ (8). The basis exchange algorithms allow one to find the minimum (4) in this case [3].

In accordance with the *K-plans* algorithm, feature vectors $\mathbf{x}_j[n]$ are divided into subsets C_k on the basis of actual central hyperplanes $H(\mathbf{w}_k[n],\theta_k)$:

$$H(\mathbf{w}_k[n],\theta_k) = \{\mathbf{x}[n]:\ \mathbf{w}_k[n]^{\mathrm{T}}\mathbf{x}[n] = \theta_k\} \tag{9}$$

where the parameters $\mathbf{w}_k[n]$ and θ_k can be found through the minimization of the convex and piecewise linear (*CPL*) criterion function $\Phi_k(\mathbf{w}[n])$ (which is described later).

The distance $\rho_H(\mathbf{x}_j[n];\mathbf{w}_k[n],\theta_k)$ of the feature vectors $\mathbf{x}_j[n]$ from the hyperplanes $H(\mathbf{w}_k[n],\theta_k)$ can be computed in accordance with the following formula:

$$\rho_H(\mathbf{x}_j;\mathbf{w}_k[n],\theta_k) = |\ \mathbf{w}_k[n]^T\mathbf{x}_j[n] / ||\mathbf{w}_k[n]|| \ - \ \theta_k]\ | / ||\mathbf{w}_k[n]|| \qquad (10)$$

The distance function $\rho_H(\mathbf{x}_j[n];\mathbf{w}_k[n],\theta_k)$ can be used for the division of feature vectors $\mathbf{x}_j[n]$ into subsets C_k in accordance with the formula (2).

3. Convex and piecewise linear (*CPL*) criterion functions $\Phi_k(\mathbf{w}[n])$

Let us consider convex and piecewise linear (*CPL*) penalty functions $\varphi_j(\mathbf{w})$ defined on the feature vectors $\mathbf{x}_j$ from the set C (1) [4]:

$$(\forall \mathbf{x}_j[n] \in C) \qquad \varphi_j(\mathbf{w}[n]) = \begin{cases} \delta - \mathbf{w}[n]^T\mathbf{x}_j[n] & \text{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] \le \delta \\ \mathbf{w}[n]^T\mathbf{x}_j[n] - \delta & \text{if} \quad \mathbf{w}[n]^T\mathbf{x}_j[n] > \delta \end{cases} \qquad (11)$$

where δ is some parameter (*margin*) ($\delta > 0$).

The penalty functions $\varphi_j(\mathbf{w})$ are equal to the absolute values $|\delta - \mathbf{w}[n]^T\mathbf{x}_j[n]|$ (Fig. 1).

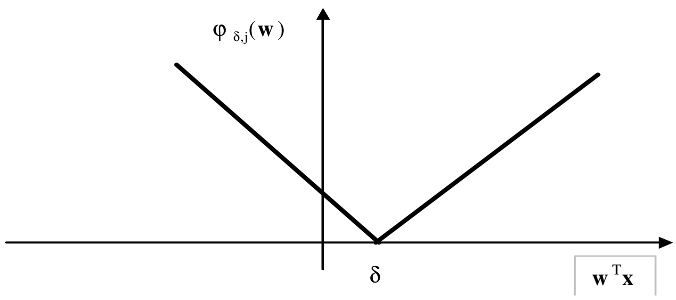

Fig. 1. The penalty function $\varphi_j(\mathbf{w}[n])$ (11)

The criterion function $\Phi_k(\mathbf{w}[n])$ is defined as the weighted sum of the penalty functions $\varphi_j(\mathbf{w}[n])$ (11) related to the vectors $\mathbf{x}_j[n]$ from the subset C_k:

$$\Phi_k(\mathbf{w}[n]) = \sum_{j \in J_k} \alpha_j\ \varphi_j(\mathbf{w}[n]), \quad \textit{where} \ \ \alpha_j > 0 \qquad (12)$$

The positive parameters α_j in the function $\Phi_k(\mathbf{w}[n])$ can be treated as the *prices* of particular vectors $\mathbf{x}_j$.

The criterion function $\Phi_k(\mathbf{w}[n])$ (12) is convex and piecewise linear as the sums of such type of functions $\alpha_j\ \varphi_j(\mathbf{w}[n])$.

Each feature vector $\mathbf{x}_j[n]$ from the set C (1) defines the hyperplane h_j in the parameter (*weight*) space R^n:

$$(\forall \mathbf{x}_j[n] \in C) \qquad h_j = \{\mathbf{w}[n]: (\mathbf{x}_j[n])^T\mathbf{w}[n] = \delta\} \qquad (13)$$

The hyperplanes h_j (13) are linked to the penalty functions $\varphi_j(\mathbf{w}[n])$ (11). The function $\varphi_j(\mathbf{w})$ (11) is equal to zero if and only if, the vector $\mathbf{w}[n]$ is situated on the hyperplane h_j (13).

Any set of n linearly independent feature vectors $\mathbf{x}_j[n]$ ($j \in J_k$) can be used for designing the non-singular matrix $\mathbf{B}_k[n] = [\mathbf{x}_{j(1)},\ldots, \mathbf{x}_{j(n)}]$ with the columns composed from these vectors. The non- singular matrix $\mathbf{B}_k[n]$ is called the k-th basis of the feature space $F[n]$. The vectors $\mathbf{x}_j[n]$ ($j \in J_k$) from this set define those n hyperplanes h_j (13) which pass through the below point (*vertex*) $\mathbf{w}_k[n]$:

$$\mathbf{B}_k[n]^T\mathbf{w}_k[n] = \boldsymbol{\delta}[n] = [\delta,\ldots,\delta]^T = \delta\ [1,\ldots,1]^T = \delta\ \mathbf{1}[n] \qquad (14)$$

or

$$\mathbf{w}_k[n] = (\mathbf{B}_k[n]^T)^{-1}\ \boldsymbol{\delta}[n] = \delta\ (\mathbf{B}_k[n]^T)^{-1}\mathbf{1}[n] \qquad (15)$$

In the case of *"short"* vectors $\mathbf{x}_j[n]$, when the number m of the vectors $\mathbf{x}_j[n]$ is much greater than the vectors dimensionality n ($m >> n$), there may exist many bases $\mathbf{B}_k[n]$ (14) and many vertices $\mathbf{w}_k[n]$ (15). It can be proved that the minimal value Φ_k^* of the criterion functions $\Phi_k(\mathbf{w}[n])$ (12) is situated in one of the vertices $\mathbf{w}_k[n]$ (15) [6]:

$$(\exists \mathbf{w}_k^*[n]) \quad (\forall \mathbf{w}[n]) \quad \Phi_k(\mathbf{w}[n]) \geq \Phi_k(\mathbf{w}_k^*[n]) = \Phi_k^* \tag{16}$$

The optimal parameter vector $\mathbf{w}_k^*[n]$ is used for the definition of the below hyperplane $H(\mathbf{w}_k^*[n], \delta)$ (9) in the feature space $F[n]$.:

$$H(\mathbf{w}_k^*[n], \delta) = \{\mathbf{x}[n]: \ \mathbf{w}_k^*[n]^T\mathbf{x}[n] = \delta\} \tag{17}$$

Theorem 1: The minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12) with $\delta \neq 0$ is equal to zero ($\Phi_k(\mathbf{w}_k^*[n]) = 0$), if and only if all the feature vectors $\mathbf{x}_j[n]$ from the subset C_k are situated on some hyperplane $H(\mathbf{w}[n],\theta))$ (9), with $\theta \neq 0$.

Proof: Let assume that the feature vectors $\mathbf{x}_i[n]$ from the subset C_k are situated on some hyperplane $H(\mathbf{w}[n],\theta)$ (9) with $\theta \neq 0$. In this case, the following equations are fulfilled:

$$(\forall \mathbf{x}_j[n] \in C) \quad \mathbf{w}[n]^T \mathbf{x}_j[n] = \theta \quad \textit{or} \quad (\delta / \theta)\, \mathbf{w}[n]^T \mathbf{x}_j[n] = \delta \tag{18}$$

thus (11)

$$(\forall \mathbf{x}_j[n] \in C_k) \quad \varphi_j((\delta / \theta)\mathbf{w}[n]) = 0 \tag{19}$$

On the other hand, if the conditions $\varphi_i(\mathbf{w}[n]) = 0$ (8) are fulfilled for all the feature vectors $\mathbf{x}_j$ from the subset C_k, then these vectors have to be situated on the hyperplane $H(\mathbf{w}[n],\theta))$ (9) .

If all the feature vectors $\mathbf{x}_j[n]$ from the subset C_k (5) are situated on some hyperplane $H(\mathbf{w}_k[n],\theta)$ (9) with $\theta = 0$, then the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12) is equal to zero ($\Phi(\mathbf{w}_k^*[n]) = 0$) only if $\delta = 0$.

It has been proved that the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12) does not depend on linear, non-singular data transformations (the *invariance property*) [4]:

$$\Phi'(\mathbf{w}_k'[n]) = \Phi(\mathbf{w}_k^*[n]) \tag{20}$$

where $\Phi'(\mathbf{w}_k'[n])$ is the minimal value (16) of the criterion functions $\Phi'(\mathbf{w}[n])$ (12) defined on the transformed feature vectors $\mathbf{x}_j'[n]$:

$$(\forall \mathbf{x}_j[n] \in C_k) \quad \mathbf{x}_j'[n] = \boldsymbol{A}[n]\, \mathbf{x}_j[n] \tag{21}$$

where $\boldsymbol{A}[n]$ is a non-singular matrix of dimension (n x n) ($\boldsymbol{A}^{-1}[n]$ exists).

The minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12) defined on the centred vectors $\mathbf{x}_i'[n] = \mathbf{x}_i[n] - \mathbf{m}_k[n]$ does not depend on translations $\mathbf{x}_j'[n] + \mathbf{b}[n]$ of the centred vectors $\mathbf{x}_j'[n]$, where $\mathbf{b}[n]$ is an arbitrary vector and $\mathbf{m}_k[n]$ is the mean vector in the subset C_k.

The minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12) is characterised by two below *montonicity properties*:

The property I (*the monotonicity in respect to reducing of the subset* C_k)

The reducing of the subset C_k to C_k' by neglecting some feature vectors $\mathbf{x}_j[n]$ can not result in an increase of the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12):

$$(C_k' \subset C_k) \Rightarrow (\Phi_k'^* \leq \Phi_k^*) \tag{22}$$

where the symbol $\Phi_k'^*$ means the minimal value (16) of the criterion function $\Phi_k'(\mathbf{w}[n])$ (12) defined on the elements $\mathbf{x}_j[n]$ of the subset C_k'.

The relation (22) can be justified by the remark that neglecting some feature vectors $\mathbf{x}_j[n]$ results in neglecting some non-negative components $\varphi_j(\mathbf{w}[n])$ (11) in the criterion function $\Phi_k(\mathbf{w}[n])$ (12).

The property II (*the monotonicity in respect to reducing of the feature space* $F[n]$)

The reduction of the feature space $F[n]$ to $F'[n']$ by neglecting some features x_i can not result in a decrease of the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12):

$$(F'[n'] \subset F[n]) \Rightarrow (\Phi_k'^* \geq \Phi_k^*) \tag{23}$$

where the symbol $\Phi_k'^*$ means the minimal value (16) of the criterion function $\Phi_k'(\mathbf{w}[n])$ (12) defined on the vectors $\mathbf{x}_j[n']$ from the feature space $F'[n']$.

The relation (23) results from the fact that the neglecting of some features x_i is equivalent to imposing an additional constraints in the form of the condition $w_i = 0$ on the parameter space R^n.

The monotonicity properties (22) and (23) constitute the basis for the proposed procedure of hidden linear dependencies extracting from data set.

4. *CPL* criterion functions $\Psi_k(\mathbf{w})$ with feature costs

Reduction of unimportant features x_i in the cost sensitive manner can be supported by the modified *CPL* criterion function $\Psi_k(\mathbf{w}[n])$ in the below form [4]:

$$\Psi_k(\mathbf{w}[n]) = \Phi_k(\mathbf{w}[n]) + \lambda \sum_{i \in I} \gamma_i \, \phi_i(\mathbf{w}[n]) \tag{24}$$

where $\Phi_k(\mathbf{w}[n])$ is given by (12), λ is the *feature cost level* ($\lambda \geq 0$), γ_i – is the *cost* of the feature x_i ($\gamma_i > 0$), $I = \{1,..., n\}$, and the *cost functions* $\phi_i(\mathbf{w})$ are defined by the unit vectors $\mathbf{e}_i[n] = [0,\ldots,1,\ldots,0]^T$:

$$(\forall i \in \{1,\ldots, n\}) \quad \phi_i(\mathbf{w}[n]) = |w_i| = \begin{cases} -(\mathbf{e}_i[n])^T\mathbf{w}[n] & \text{if } (\mathbf{e}_i[n])^T\mathbf{w}[n] < 0 \\ (\mathbf{e}_i[n])^T\mathbf{w}[n] & \text{if } (\mathbf{e}_i[n])^T\mathbf{w}[n] \geq 0 \end{cases} \tag{25}$$

The criterion function $\Psi_k(\mathbf{w}[n])$ (24) is the convex and piecewise linear (*CPL*) as the sum of the *CPL* functions $\Phi(\mathbf{w}[n])$ (12) and $\lambda\ \gamma_i\ \phi_i(\mathbf{w}[n])$ (25). The optimal point $\mathbf{w}_\lambda'[n]$ constitutes the minimal value of the criterion function $\Psi_k(\mathbf{w}[n])$:

$$(\exists \mathbf{w}_\lambda'[n]) \ (\forall \mathbf{w}[n]) \quad \Psi_k(\mathbf{w}[n]) \geq \Psi_k(\mathbf{w}_\lambda'[n]) \tag{26}$$

Each *CPL* cost function $\phi_i(\mathbf{w}[n])$ tends to reach the condition $w_i = 0$ (24) through the minimization of the function $\Psi_k(\mathbf{w}[n])$ (26) and to reducing the feature x_i. The influence of the cost functions $\phi_i(\mathbf{w}[n])$ increases with the value of the parameter λ . The increase of the cost level λ can lead to reducing additional features x_i.

Each unit vector $\mathbf{e}_i[n]$ defines the below hyperplane $h_{0,i}$ in the parameter space R^n:

$$(\forall i \in \{1,2,\ldots, n\}) \qquad h_{0,I} = \{\mathbf{w}[n]: (\mathbf{e}_i[n])^T\mathbf{w}[n] = 0\} \tag{27}$$

The minimum point $\mathbf{w}_\lambda'[n]$ (26) of the function $\Psi_k(\mathbf{w}[n])$ (24) is situated in one of the vertices $\mathbf{w}_k'[n]$ ($\mathbf{w}_\lambda'[n] = \mathbf{w}_k'[n]$) defined by the equation of the below type (14):

$$\mathbf{B}_k[n]^T\, \mathbf{w}_k'[n] \;=\; \boldsymbol{\delta}'[n] \;=\; [\delta,\ldots, \delta, 0,\ldots,0]^T \tag{28}$$

In this case, the columns of the matrix $\mathbf{B}_k[n]$ can be composed partly of some feature vectors $\mathbf{x}_j[n]$ and partly of some unit vectors $\mathbf{e}_i[n]$. The vertex $\mathbf{w}_k'[n]$ (15) is the point of intersection of hyperplanes h_j (13) defined by some feature vectors $\mathbf{x}_j[n]$ and hyperplanes $h_{0,i}$ (27) defined by unit vectors $\mathbf{e}_i[n]$. The minimum point $\mathbf{w}_\lambda'[n]$ (26) of the function $\Psi_\lambda(\mathbf{w}[n])$ (24) is situated in one of such vertices $\mathbf{w}_k'[n]$, which is the intersection point of n hyperplanes h_j (13) and $h_{0,i}$ (27).

The features x_i, which are linked to the unit vectors $\mathbf{e}_i[n]$ in the optimal basis $\mathbf{B}_k^*[n]$ (28) fulfil the below equation and can be reduced without changing of the minimal value $\Psi_k(\mathbf{w}_\lambda'[n])$ (26):

$$(\mathbf{e}_i[n])^T\, \mathbf{w}_k'[n] = 0 \;\Rightarrow\; w_{ki} = 0 \Rightarrow \text{feature } x_i \text{ is reduced} \tag{29}$$

where $\mathbf{e}_i[n] = [0,\ldots,0,1,0,\ldots,0]^T$ is the i-th unit vector, and $\mathbf{w}_k'[n] = [w_{k1},\ldots\ldots,w_{kn}]$.

The number of the reduced features x_i can be increased by an increasing the feature cost level λ in the criterion function $\Psi_\lambda(\mathbf{w}[n])$ (24).

5. Extracting of hidden linear dependencies

The procedure of hidden linear dependencies extracting from data set C (1) can be based on the *CPL* criterion functions $\Phi_k(\mathbf{w}[n])$ (12) and $\Psi_k(\mathbf{w}[n])$ (24) The monotonicity properties (22) and (23) are particularly important in the proposed procedure.These monotonicity properties (22) and (23) are valid not only for the minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the function $\Phi_k(\mathbf{w}[n])$ (12) but also for the minimal value $\Psi_k(\mathbf{w}_\lambda^*[n])$ (26) of the function $\Psi_k(\mathbf{w}[n])$ (24).

The multistage procedure of extracting of hidden linear dependency is described by the below successive steps:

i. Two small, positive parameters (*margins of precision*) τ_1 and τ_2 are defined ($\tau_2 \geq \tau_2 > 0$), the value $k = 1$ and the initial set $C_k' = C$ (1) of all the feature vectors $\mathbf{x}_i[n]$ are fixed

ii. There is the computed minimal value $\Phi_k(\mathbf{w}_k^*[n])$ (16) of the criterion function $\Phi_k(\mathbf{w}[n])$ (12). The function $\Phi_k(\mathbf{w}[n])$ is defined on all the feature vectors $\mathbf{x}_j[n]$ from the set C_k'.

iii. The minimal number of the feature vectors $\mathbf{x}_i[n]$ is omitted from the set C_k' in order to reach the condition $\Phi_k(\mathbf{w}_k^*[n]) \leq \tau_1$. Such vectors $\mathbf{x}_j[n]$ are reduced which caused the smallest increase of the value $\Phi_k(\mathbf{w}_k^*[n])$ (16). The remaining feature vectors $\mathbf{x}_j[n]$ form the k-th *linearly dependent cluster* C_k.

iv. The maximal number of the features x_i is omitted from the feature space $F[n]$, while preserving the condition $\Phi_k'(\mathbf{w}_k'[n']) \leq \tau_2$ in the new feature subspace $F'[n']$ ($F'[n'] \subset F[n]$). The vector $\mathbf{w}_k'[n']$ constitute the minimum (26) of the function $\Psi_k(\mathbf{w}[n])$ (24). The dimensionality of the feature vectors $\mathbf{x}_j[n]$ from the cluster C_k is reduced from n to n' by a successive increase of the *feature cost level* λ. (24).

v. The below linear relation between features x_i from the feature subspace $F'[n']$ ($x_i \in F'[n']$) is formed on this basis

$$w_{\lambda,i(1)}'\, x_{j,i(1)} + \ldots\ldots + w_{\lambda,i(n')}'\, x_{j,i(n')} = \delta \tag{30}$$

where $\mathbf{x}_i[n'] = [x_{i,i(1)},..,x_{i,i(n')}]^T$ is the feature vector ($\mathbf{x}_j[n'] \in F'[n']$) reduced during the previous stage, and $\mathbf{w}_\lambda'[n'] = [w_{\lambda,i(1)}',\ldots, w_{\lambda,i(n')}']^T$ is the optimal vector (26) with all the components $w_{\lambda,i}'$ different from zero ($w_{\lambda,i}'$,0).

vi. The set C_k' is reduced by neglecting such feature vectors $\mathbf{x}_j[n]$ which constitute linearly dependent cluster C_k. If the set C_k' is not empty ($C_k' \neq \emptyset$), then the value of the parameter k is increased by one ($k \rightarrow k+1$) and the next stage is started from the step *ii.*

The above procedure allows to extract K linearly dependent clusters C_k from the data set C (1). Each cluster C_k is represented by K linear relations (30). As opposed to the *K- means* algorithm, the number K is not fixed at the beginning of this procedure. The number K of the clusters C_k reflects the structure of the data set C (1). Let us remark that the relation (30) allows to form n' regression type models. Each component $x_{i,i(k)}$ can represent dependent variable (feature) $x_{i(k)}$ and the remaining n' - 1 components $x_{j,i(k')}$ can represent dependent variables.

Such regression type models have local properties. This means that each model represents feature vectors $\mathbf{x}_j[n]$ from one particular cluster C_k.

6. Concluding remarks

The problem of extracting linearly dependent patterns from data sets is considered in the paper. The proposed approach is based on the minimization of two convex and piecewise linear (*CPL*) criterion functions $\Phi_k(\mathbf{w}[n])$ (12) and $\Psi_k(\mathbf{w}[n])$ (24̃).

Extraction of hidden, linearly dependent patterns is considered here as a problem of cluster analysis. The *K-plans* algorithm, similarly to the *K-means* algorithm has been proposed as one of the tools for solving this problem. In this approach each linearly dependent cluster C_k is represented by some central hyperplane $H(\mathbf{w}_k[n],\theta_k)$ (9)

Alternative approach is based on what is described in Paragraph 5 as a sequence of the functions $\Phi_k(\mathbf{w}[n])$ (12) and $\Psi_k(\mathbf{w}[n])$ (24̃) minimizations. In this approach there is no need to fix the number K of clusters C_k beforehand. The extraction of linearly dependent clusters C_k is linked here to designing the regression type models (30). As a result, the data set C (1) is represented by the family of K linear models (30). Such representation allows to expose the internal linear structure hidden in the set C (1).

The usefulness of the extracted linearly dependent patterns and models should be verified in many ways. In accordance with data mining or exploratory analysis standards, experts in the fields should have the final judgments concerning the extracted patterns.

Bibliography

[1]. Hand D., Smyt P., Mannila H.: *Principles of Data Mining*, MIT Press, Cambridge, MA 2001

[2]. Duda O. R, Hart P. E., Stork D. G.: *Pattern Classification,* J. Wiley, New York, 2001.

[3]. Bobrowski L., Bezdek J. C., "*C-means clustering with the L_1 and L_∞ norms*", *IEEE Transactions on Systems Man and Cybernetics,* Vol. 21, No. 3, pp. 545-554, 1991.

[4]. Bobrowski L.: "*CPL clustering with feature costs*", *ICDM*2008, Leipzig, Germany

[5]. Bobrowski L.: "*Design of piecewise linear classifiers from formal neurons by some basis exchange technique*" *Pattern Recognition*, **24**(9), pp. 863-870, 1991

[6]. Bobrowski L.: *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych* (*Data mining based on convex and piecewise linear (CPL) criterion functions*) (in Polish), Technical University Białystok, 2005